Steering Towards Sustainability: Predictive Approaches to Lowering HC Emissions at Ports

Tanvi S. Patil

San Diego State University

MIS 790 Culminating Experience

Carlos D. Paternina-Arboleda

April 7th, 2024

# Abstract

As global trade and maritime transport demand have risen, so too have concerns over pollutant emissions from ports, crucial junctures in the maritime industry contributing significantly to environmental degradation. This report delves into predictive modeling of hydrocarbon (HC) emissions from ships at ports, aiming to identify key factors influencing emissions and explore effective reduction strategies. Through the evaluation of various predictive models, including Ridge Regression and XGBoost, this study highlights the impact of operational practices and fuel usage on emission levels. Findings suggest that targeted technological and operational interventions can markedly reduce both emissions and operational costs. By bridging the gap between scientific research and practical application, this report aims to inspire port authorities and maritime stakeholders to embrace innovative technologies and strategies for enhancing environmental sustainability. It posits that through informed decision-making and collaborative efforts, it is possible to significantly improve the environmental performance of maritime transport, paving the way for a more sustainable future for the industry.

*Keywords*: HC Emissions, Predictive Modelling, Linear Regression, Polynomial Regression, Ridge Regression, XGBoost, Interaction effects,

Steering Towards Sustainability: Predictive Approaches to Lowering HC Emissions at Ports

# 

**Introduction**

The Marine Industry represents the main part of world trade, as it involves transporting products from one location to another. Shipping activities are viewed as a driving force behind the world's economies, but the environmental effect of ship emissions is treated delicately. HC(hydrocarbons) from ships, in turn, are a matter of special attention because they can get into the air and thus influence air pollution and climate. VOCs, which are primary components of hydrocarbon compounds, are attributed to the release of ground-level ozone and secondary organic aerosols with their implications on human health and air quality, both short- and long-term. (*Volatile Organic Compounds’ Impact on Indoor Air Quality | US EPA*, 2023b)

According to Deniz et al. (2010), The main part is ships contribute the air pollution making gas components that are important for the atmosphere, such as nitrogen oxide (NOx), sulfur oxide gases (SO2), carbon oxides, volatile hydrocarbons (HC), and particulate matter.

However, these emissions have serious consequences on human health and the environment, they threaten human health through respiratory cancers, impaired lung functions in the eyes of children, chronic bronchitis, worsening of asthma hospitalization rise of respiratory and cardiovascular diseases. Moreover, pollutants affect acid rain, plants, water, visibility, and consequently, global warming (Manisalidis et al., 2020). The emissions have a significantly distractive effect in the protected waters, narrow seas, canals, channels, and estuary districts where the activity of about 80% of the worldwide fleet is either at anchorage or operating very close to the coast. According to an estimate, ship emissions in port regions contribute up to 55%-77% to the total emissions inventory, whilst the ship sources and contributions to the emission inventory in densely populated land areas create a range from 5 to 30%. These releases bring about health impacts, and other bigger environmental problems, and therefore they should be dealt with and reduced in quantity level (Lee et al., 2021).

The dataset under consideration provides a detailed record of HC emissions from ships at a port, including metrics such as ship operations, fuel consumption, and emission levels. Among the other elements that contribute to this are, the ship's type, load, fuel type, engine speed, and docking and sailing times. The analysis's goal is to find significant predictors of HC emissions and understand how different operational features of ships affect emission levels.

The primary goal of this analysis is to get an understanding of the elements that have a substantial impact on HC emissions in the port environment. Understanding these aspects is critical for creating successful methods to reduce emissions and improve environmental sustainability in maritime activities. Furthermore, this investigation aims to create prediction models utilizing machine learning approaches to estimate HC emissions based on ship features and activities. Such models can be useful for planning and implementing emission-reduction strategies.

Specific questions and hypotheses explored in this analysis include:

What are the main predictors of HC emissions from ships, and how do they interact to influence overall emission levels? It is hypothesized that fuel type, engine speed, and ship type are significant predictors of HC emissions. It is anticipated that emission levels vary significantly between different stages of ship operations.

Can machine learning models effectively predict HC emissions based on available data?

Is there a significant difference in levels between ships using different types of fuel? It is hypothesized that ships using heavier fuels will have higher HC emissions compared to those using cleaner fuels.

It is anticipated that the results of this analysis will add to the body of knowledge already available on ship emissions and offer insightful information that will help shipping firms, port authorities, and legislators develop focused emission reduction plans. By identifying key factors influencing HC emissions and developing accurate predictive models, this analysis aims to support the maritime industry's transition towards more sustainable and environmentally friendly practices.

**Literature review**

Maritime transport is an essential part of world trade, accounting for more than 80% of the international trade volume carried by ships. However, this industry significantly contributes to environmental degradation, particularly through the emissions of hydrocarbons (HC), which negatively impact human health, air quality, and climate change. With a focus on applicable literature on port emissions, viable solutions, relevant technology, and previous study approaches, this review attempts to give a thorough overview of the latest research on greenhouse gas emissions from ships at ports.

Ships contribute significantly to air pollution by emitting large amounts of hydrocarbons (HC), particulate matter (PM), sulfur oxides (SOx), and nitrogen oxides (NOx), at local and regional levels around lakes, rivers, and seas (Yeh et al., 2022; Wu et al., 2023). In 2011, emissions from ships at ports included 18 million tonnes of carbon dioxide (CO2), 0.4 million tonnes of NOx, 0.2 million tonnes of SOx, and 0.03 million tonnes of PM10. Notably, tankers and containerships accounted for approximately 85% of these emissions (Merk & International Transport Forum, 2014).

More than 2.5 million tons of carbon dioxide equivalent (CO2e) and other hazardous pollutants that are harmful to human health, like hydrocarbons (HC), nitrogen oxides (NOx), sulfur oxides (SOx), and particulate matter (PM), were released in 2019 from the three biggest ports in the United States: The Ports of Los Angeles, Long Beach, New York, and New Jersey. At certain ports, progress has been achieved in lowering emissions through sustainability initiatives such as converting to fuels with lower sulfur emissions, shore powering—providing ships berthed at the port with electricity generated on land—lowering vessel speeds, and making energy-efficient investments (Environmental and Energy Study Institute (EESI), n.d.).

Volatile organic compounds (VOCs), which include methane (CH4), ethane, and various other hydrocarbons, are produced by HC emissions. These VOCs aid in the production of secondary organic aerosols and ground-level ozone, further exacerbating environmental and health issues. The shipping industry is responsible for a considerable share of greenhouse gas emissions, with ships' CO2 emissions estimated to constitute about 2.89% of global anthropogenic emissions in 2018 (Aakko-Saksa et al., 2023).

To reduce port-related emissions various strategies were discussed by (U.S. Environmental Protection Agency, 2016), for hydrocarbon (HC) emissions including replacing older diesel fleets with cleaner technology, implementing operational improvements like shore power to cut down on emissions from idling and convert to greener fuels such as low-sulfur diesel. Additionally, adopting zero-emissions technologies, such as electric cargo handling equipment and locomotives, can significantly reduce emissions. Improving fuel efficiency in rail locomotives and harbor craft also contributes to emission reductions (Ports Initiative | US EPA, 2024). Moreover, powering ships with electric engines can reduce fuel consumption, SO2, CO2, and NOx and HC emissions (Barberi et al., 2021b).

Efforts to calculate and minimize CO2 emissions in ports are increasing, with some ports initiating carbon footprint reporting. However, the absence of standardized methodologies hampers accurate CO2 quantification and comparison across ports. International initiatives like the World Ports Climate Initiative (WPCI) and the Clean Cargo Working Group (CCWG) aim to raise awareness and provide tools for calculating CO2 footprints, highlighting the global push for sustainable maritime transport. (Azarkamand et al., 2020).

Besides the solutions discussed earlier in the work, the literature also provides information regarding the other strategies for reducing HC ship emission at ports, apart from reducing the sulfur (SO (2)). One of the major solutions is the switch from heavy fuel oil (HFO) to ultra-low-sulfur fuel oil (ULSFO). As reported by Yeh et al. (2022) in their study, conversion from HFO to ULSFO leads to a reduction in the releases of pollutants of particulate matter and SOx by almost 96% and NOx by 17%, respectively. On the other hand, a decrease of 16.9% and 36.1% in CO and HC emissions, respectively, has been observed with ULSFO indicating incomplete combustion. This emphasizes the significance of fuel consumption efficiency optimization from engine combustion generated when using low-sulfur fuel.

A real-time analysis of energy use and air pollution emissions during a transpacific container ship voyage is also given by Yeh et al. (2022). The results recommend more investigation into energy-efficient reefers, optimizing generator engines, and implementing shore power.

The Port of San Diego has adopted an all-inclusive CAP (Climate Action Plan) to effectively tackle those challenges and adhere to the state's goals. The CAP involves the adoption of policies and the provision of actions pertaining to waste reduction in conjunction with recycling, renewable energy, water conservation and recycling, energy efficiency and conservation, transportation, and land use planning. The plan intends to meet its target of cutting 10% GHG emissions by 2020 compared to the level in 2006 (855,489 vs 745,695 metric tons of CO2 equivalent) as well as the 25% reduction goal by 2035. The Port's GHG emissions mainly result from the movements of goods and energy consumption within the transportation sector, and it is predicted that future emissions will heighten as a result of new commercial development initiatives and anticipated rises in cruise and cargo traffic. ((Port of San Diego, 2013), Wright, P. (2013))

The Port of Volos in Greece has implemented several steps taken to lower emissions, including the use of real-time online systems to optimize truck load factors, increasing the number of heavy goods vehicles (HGVs) powered by alternative fuels, and implementing intelligent transport systems for local traffic management. These measures aim to improve traffic flow, reduce fuel consumption, and lower emissions from diesel-powered HGVs (Karakikes et al., 2018).

The Port of Barcelona in Spain has undertaken green initiatives across various categories, such as investing in infrastructure to support alternative fuels like LNG, providing electric connections for marine vessels, and offering incentives to shipping companies for environmental performance. The port has also concentrated on electrifying and gasifying port terminal machinery, replacing diesel-powered land vehicles coupled with natural gas or electric power vehicles, and improving rail and short-sea shipping facilities to minimize road transit. (Gonzalez-Aregall & Bergqvist, 2020).

In the study of hydrocarbon (HC) emissions in ports, a systematic review methodology was applied to analyze and identify relevant methodologies for estimating ship emissions. The review followed PRISMA guidelines using a bottom-up, multi-layer analytical method (Zhang et al., 2017), focusing on both methods that are top-down and bottom-up for inventory development. The top-down method, also known as the fuel-based method, employs statistics on fuel sales and emission factors to estimate total emissions. The bottom-up technique, which is recommended for comprehensive information on ship movements, integrates fuel or engine energy production with time values and emission parameters related to particular tasks like navigating, maneuvering, and hoteling. (Bojić et al., 2022).

The practical methods to stop ship emissions in ports include using shore electricity, sometimes referred to as cold ironing. By giving ships at anchor electrical power, shore power systems enable them to turn off their engines and cut back on NOx, PM, SOx, and HC emissions. Studies have demonstrated that shore power systems can greatly lower emissions; Estimates suggest that when a quarter to two-thirds of all ships calling at US ports use shore power, the benefits to air quality might reach 70–150 million US dollars annually. Wan et al. (2021).

The adoption of carbon-neutral and low-carbon fuels, such as biofuels, e-fuels, and liquefied biogas (LBG), has been proposed to address GHG emissions. Technologies for reducing emissions from exhaust, such as selective catalytic reduction (SCR) and exhaust gas recirculation (EGR), have been highlighted as effective methods for reducing harmful pollutants like NOx and SOx (Aakko-Saksa et al., 2023). The implementation of emission control areas (ECAs) requires large ocean-going vessels to use lower sulfur fuel, which has almost 90% decreased fuel-based PM emissions (Aeroqual, 2021).

A subsequent study by Flowers et al. (2002) employed multi-zone simulations to forecast emissions of hydrocarbons and carbon monoxide in the combustion of isooctane homogeneous charge compression ignition (HCCI) engines. Their results showed that The model's estimation of emissions of carbon monoxide and hydrocarbons was impacted by how many zones were used in the simulation.

Fu et al. (2022) present a unique approach to estimating ship emissions by combining real-time emission data with factors influencing emissions. This method accounts for spatial and temporal variations in shipping activities, leading to more accurate emission estimates. The results demonstrate the potential of this approach for improved environmental management and decision-making in the maritime sector.

Xiao et al. (2023) use multiscale spatially weighted regression to assess port pollutant emission characteristics in the United States. The study makes use of Automatic Identification System (AIS) data and analyzes it taking into account factors like population, throughput, and coastal length. The results provide insights into the spatial distribution and influencing factors of port emissions, contributing to a better understanding of emission patterns and potential mitigation strategies.

In their research, Darbra et al. (2004b) have discussed the Self-Diagnosis Method. SDM is a tool made to evaluate seaports' environmental management, helping ports review their management activities and identify improvement opportunities. Based on the ISO 14001 structure, it serves as a "first-level" tool for non-expert users and supports port managers in reviewing environmental performance, comparing it against European benchmarks, and moving towards implementing an environmental management system.

A range of mitigation and monitoring techniques for ship emissions are presented in the literature; these are essential for enhancing port air quality and minimizing environmental damage. To reduce the amounts of sulfur oxides (SOx), Hydrocarbons (HCs), and other pollutants, one strategy is to employ exhaust gas cleaning devices, such as scrubbers, and low-sulfur fuels. Differential Optical Absorption Spectroscopy (DOAS), one of the remote sensing methods that has been used to monitor ship emissions in real-time, has the advantages of cost-effectiveness, environmental adaptability, and coverage (Wu et al., 2023). These technologies offer useful information for evaluating how well emission reduction strategies are working.

In recent years, the quest for sustainable port operations has gained momentum, with a focus on mitigating emissions and pollution. Bjerkan and Seter (2019) highlight the critical role of environmental management systems in effectively monitoring port activities. As a vital tactic to lower greenhouse gas emissions and improve air quality in port areas, they support the use of alternative fuels like biofuels and liquefied natural gas (LNG).

The potential for modal shifts in transportation and vessel speed reductions to significantly reduce the carbon footprint of port operations is heavily stressed. Accurate emissions inventories are mostly dependent on the use of machine learning methods, such as multiple linear regression (MLR), support vector regression (SVR), and artificial neural networks (ANN), for fuel consumption and emission prediction. These models, trained on various features including vessel characteristics and engine parameters, underline the importance of empirical research in validating their effectiveness and supporting decision-making processes in ports aiming for sustainability (Cammin et al., 2022).

Yao et al. (2022) present a novel integrated model for estimating carbon emissions in the case study of the Port of Los Angeles. It combines the long short-term memory (LSTM), autoregressive integrated moving average with explanatory variable (ARIMAX), and stochastic impacts by regression on population affluence and technology (STIRPAT). This model demonstrates significant improvement in estimation accuracy compared to existing techniques. The study shows a significant relationship between carbon emissions and port throughput, indicating that managing carbon emissions in container ports requires a methodical approach that combines strategic operational optimization with predictions based on historical data.

The paper by Paternina-Arboleda et al. (2023) emphasizes the significance of ports in global trade, as well as the fact that their emissions have a detrimental effect on air quality and climate change. Based on machine learning frameworks, the research suggests how much SO2 a ship emits during various port operations like anchoring, maneuvering, and cruising. The AutoML TPOT framework is the most prominent model among those analyzed for their ability to estimate SO2 emissions. The study shows the crucial role of empirical studies in verifying the implications of such models. Also, it aids the decision-making processes that port authorities consider when on the way to achieving sustainability (Paternina-Arboleda et al., 2023).

The paper by Hussaın et al. (2022) concentrates on PM exposure among port workers from ports operations is the area of study. The investigation underlines the fact that port activities cause damage to the air and the health dangers of being exposed to PMs are elevated for port workers. Research mobilizes modern monitoring tools and models to quantify the level of exposure and accentuates the importance of imposing effective pollution control measures in port areas. This research justifies the need to maintain more air quality monitoring and health surveillance, to safeguard port workers' health (He et al., 2023).

Achieving green ports requires a comprehensive approach that involves the integration of advanced technologies, data-driven models, and regulatory measures. It is also crucial to foster collaboration between researchers and port authorities to bridge the gap between research and practice, ensuring that research findings are applicable and beneficial for practical decision-making (Bjerkan & Seter, 2019; Cammin et al., 2022; Yao et al., 2022).

**Data Description**

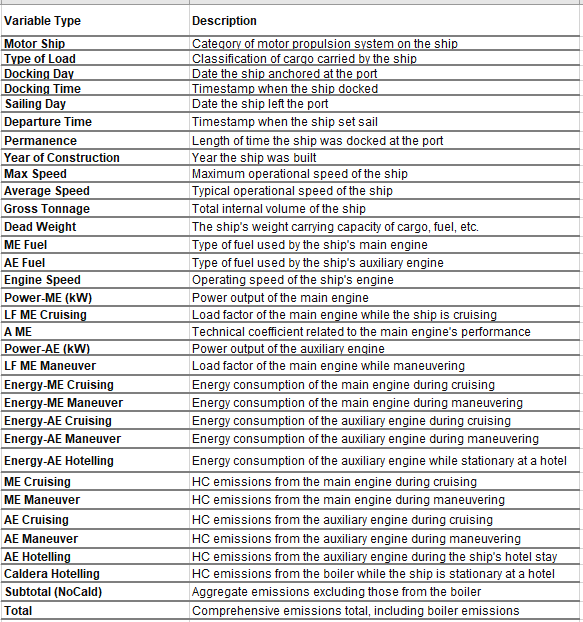
The dataset under analysis was obtained from a port authority database, which records emissions from ships during their docking and sailing activities. The data encompasses various parameters related to ship operations and environmental impact, specifically focusing on hydrocarbon (HC) emissions.

The dataset contains 651 observations (rows) and 38 variables (columns) that cover a variety of variables such as ship kinds, fuel usage, engine speed, and pollution levels.

The description of each variable in the dataset is provided here.

**Figure 1**

*Description of all the variables in the dataset*



**Data Preprocessing**

*Handle missing values:*

In the data preprocessing phase of the script, missing values are identified and addressed to ensure the dataset's integrity. The df.isna().sum() function was used to count missing values in each column, revealing that the 'Year\_of\_Construct' and 'Dead\_Weight' columns have 1 and 43 missing values, respectively, while the 'IMO' column also has 43 missing values. To handle these missing values, the df = df.dropna() command was employed to remove any rows containing missing values. This approach ensures that the dataset used for model training is complete, without any incomplete records. By taking these steps to address missing values, we prepared the dataset for further analysis and model training, ensuring that the data is clean and reliable.

*Encode categorical variables*

In the data preprocessing stage, categorical variables were converted into numerical formats through label encoding, a process essential for machine learning algorithms that require numerical input. Specifically, the script applies label encoding to the 'Motor\_Ship', 'Type\_of\_Load', 'EM\_Fuel', 'AE\_Fuel', and 'Engine\_Speed' columns. Each category in these columns is given a distinct integer by this encoding, which turns them into numerical labels. This step is crucial for preparing the dataset for subsequent analysis and modeling, as it ensures that all features are in a format compatible with machine learning algorithms.

**Feature Selection**

*Identify relevant features*

To identify the important and relevant features correlation matrix was plotted. An essential analytical tool for analyzing the connections between each measured variable in the dataset is the correlation matrix. The correlation coefficient between two variables, which expresses the magnitude and direction of their linear relationship, is represented by each element in this matrix. A significant positive or negative correlation is shown by values near 1 or -1, respectively, implying a direct or inverse proportionality in the variability of one variable in relation to the other. Conversely, coefficients near zero imply a negligible linear relationship. For feature selection in predictive modeling, this matrix is invaluable. Features that exhibit a substantial correlation with the target variable are often considered good predictors and hence, prime candidates for inclusion in the model.

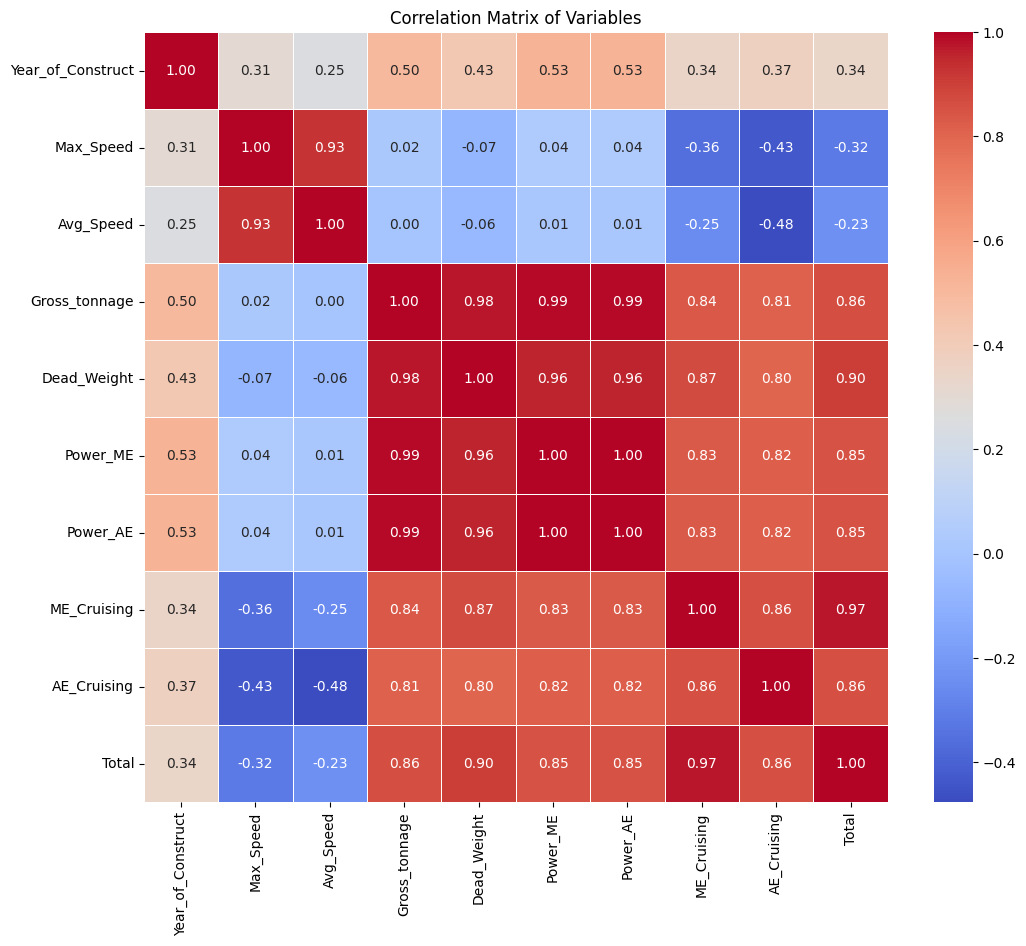
In the analysis of the correlation plot provided for the dataset in question, certain patterns are discernible that inform the feature selection process for our predictive modeling endeavors. For instance, the variables 'Gross\_tonnage', 'Dead\_Weight', 'Power\_ME', and 'Power\_AE' show a notably high correlation with the target variable 'Total'. This strong relationship indicates that as the values for tonnage, dead weight, and engine power change, there is a corresponding significant impact on the total emissions, making these variables potentially powerful predictors in the model.

Moreover, there is an observable trend of multicollinearity among some of the independent variables. Notably, 'Gross\_tonnage' correlates highly with 'Dead\_Weight', and 'Power\_ME' with 'Power\_AE'. This strong interdependency suggests that these pairs of variables share a significant amount of information. In predictive modeling, such multicollinearity can be problematic, potentially leading to inflated variances for the parameter estimates, which can undermine the statistical significance of the predictors. Therefore, with careful consideration, we have to choose to either combine these variables into a single feature or select the one that provides the most unique and relevant information for the model.

The lower correlations present between variables such as 'Year\_of\_Construct' with 'Max\_Speed' and 'Avg\_Speed', although not negligible, suggest a weaker predictive relationship with the total emissions. These variables may still provide value to the model, albeit to a lesser extent, and their inclusion would depend on the specific modeling goals and the desired complexity of the final model.

**Figure 2**

*Correlation matrix of variables*

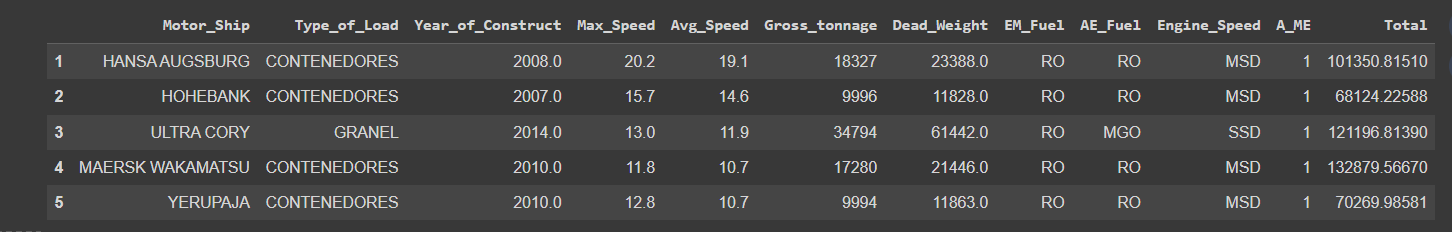


*Description of Relevant Features:*

In the context of modeling hydrocarbon emissions from ships at ports, understanding and selecting relevant features from the data is critical. Here we describe key features that are indicative of the factors influencing emissions and are, therefore, integral to the model's predictive power

**Figure 3**

*Relevant features used for further predictions*



**Motor\_Ship:** This categorical feature indicates the type of motor a ship is equipped with.

**Type\_of\_Load:** The type of cargo a ship carries can affect the vessel's stability and energy usage, potentially influencing emissions.

**Year\_of\_Construct:** Represents the year the ship was built

**Max\_Speed and Avg\_Speed:** These continuous variables represent the maximum and average speed of the ships, respectively.

**Gross\_tonnage**: This feature refers to the overall internal volume of a ship and is a proxy for the ship's size and cargo capacity.

**Dead\_Weight:** Denoting the total weight a ship can safely carry, this measure includes cargo, fuel, passengers, and crew.

**EM\_Fuel and AE\_Fuel:** These features identify the kind of fuel that the vessel’s main engine is using (EM) and auxiliary engine (AE).

**Engine\_Speed:** This indicates how fast the ship's engines are running, which is linked to the rate of fuel consumption.

**A\_ME:** Technical coefficient related to the main engine's performance

**Total:** Represents the total HC emissions produced by the ship.

Collectively, these features provide a comprehensive dataset that reflects various aspects of shipping operations that are significant contributors to HC emissions. Through machine learning algorithms, these features are analyzed to understand their influence on emissions, thereby enabling the development of models that can accurately predict emission levels based on ship and voyage characteristics. This information is crucial for the formulation of strategies aimed at reducing the environmental impact of port activities.

**Method Selection**

In assessing the predictive capabilities for hydrocarbon (HC) emissions in maritime environments, the study's methodological strategy was both rigorous and multifaceted. Selecting the most appropriate predictive algorithms was paramount, focusing on those well-suited to handle the complexities of emissions data.

**Figure 4**

*Graphical depiction of the method*

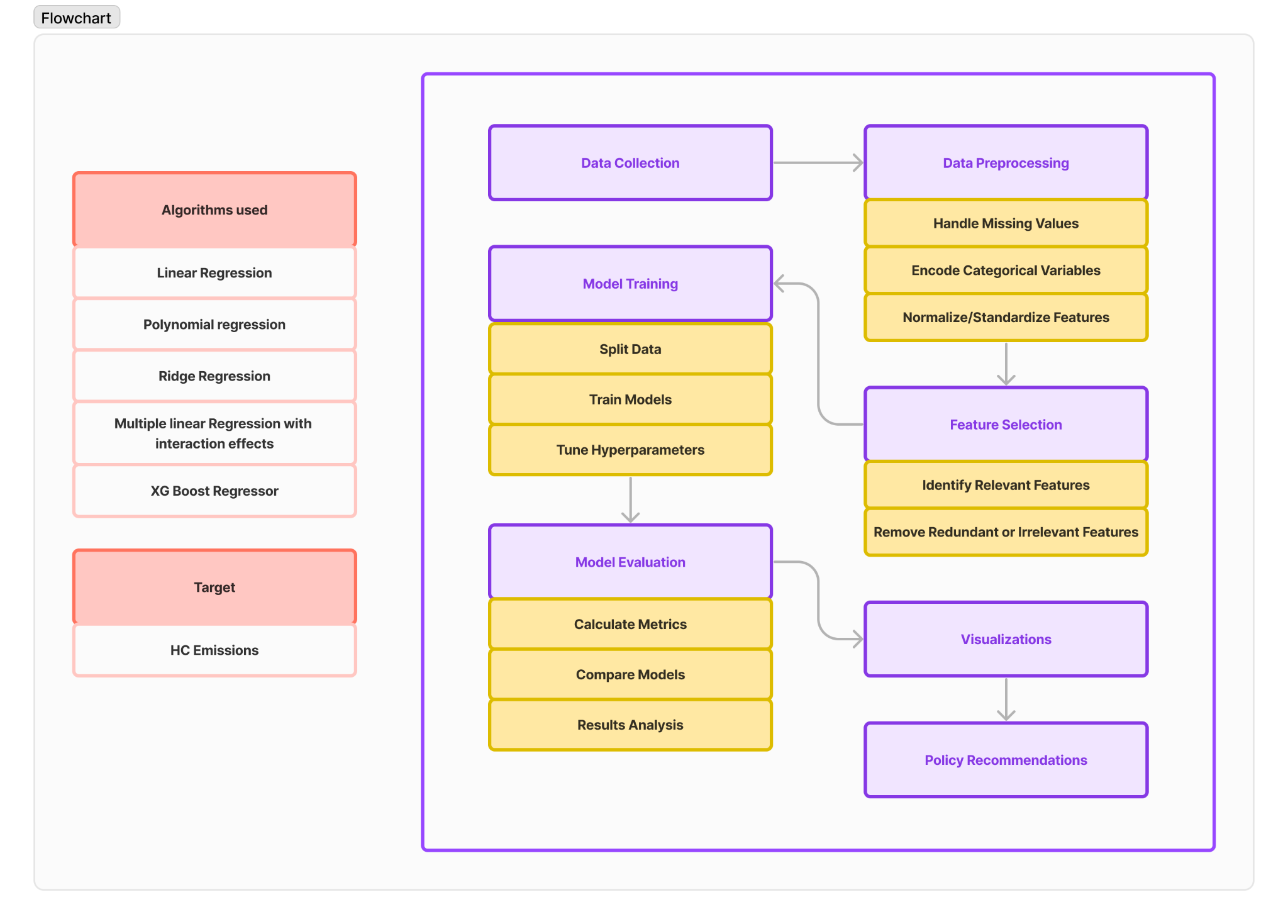


Figure 4 shows the graphical description of the method selection. To build the prediction model and assess its effectiveness, the dataset is divided into training and test sets. The model is trained using the training set, which consists of 75% of the data, so that it can understand how the features relate to the target. The remaining 25% of the data is reserved as the test set, which will not be seen by the model during training. This test set acts as new data for the model, providing an unbiased evaluation of its performance.

Following data preparation and feature selection, the Model Initialization phase saw the deployment of several models, each with unique characteristics and assumptions about the data. Linear Regression was employed as a baseline model, assuming a straightforward linear relationship between the target and the input features. To capture more complex, non-linear relationships, Polynomial Regression was introduced, enhancing the linear model with polynomial terms. Ridge Regression was also utilized, incorporating L2 regularization to penalize large coefficients, thus preventing overfitting and enhancing model generalizability. Multiple Linear Regression with Interaction Effects took this further by not only considering polynomial terms but also focusing on the interactions between features, aiming to capture the combined effects of variables on emissions. The XGBoost Regressor, known for its efficiency and effectiveness in handling complex datasets, was another critical model deployed, leveraging gradient boosting techniques to tackle the predictive challenge.

Feature Transformation and Scaling played a vital role in preparing the data for modeling. Polynomial Features were generated to allow models to understand and utilize non-linear relationships within the data. Standard Scaling was applied to normalize the feature set, ensuring that all variables contributed equally to the model training process, which is particularly crucial for models like Ridge Regression and when dealing with polynomial features.

In the Model Training and Hyperparameter Tuning stage, each model was meticulously trained on the training set. For the XGBoost model, GridSearchCV was employed for hyperparameter tuning, systematically searching through a predefined grid of hyperparameters to find the combination that yielded the best performance, as measured by cross-validation.

In order to evaluate each model, emissions predictions were generated on the test set, and important metrics including Mean Squared Error (MSE), R-squared (R²), Mean Absolute Error (MAE), and Explained Variance were computed. These metrics gave each model's accuracy, fit, and capacity to generalize to new data a numerical evaluation.

Cross-validation provided an enhanced validation technique by evaluating the model's performance across several data subsets, especially for models like XGBoost and Ridge Regression. This approach ensured that the model's predictive power was not merely a function of a specific train-test split but held more broadly across the dataset.

Finally, the Optimization and Selection phase involved identifying the best parameters for models undergoing hyperparameter tuning, like XGBoost. The performance of all models was then evaluated and compared, allowing for an informed selection of the model(s) that offered the best balance between accuracy, complexity, and interpretability for predicting HC emissions.

**Description of Machine Learning Tools**

***Linear Regression:***

Predictive modeling uses a basic machine learning tool called linear regression. This approach is especially appropriate for situations in which the dependent variable (outcome) and the independent variables (predictors) have a linear relationship.

Objective: Finding a linear relationship between one or more independent variables (X) and a continuous dependent variable (Y) is the main objective of linear regression. This relationship is represented by a linear equation:

=

Where,

ŷ is the predicted value of a dependent variable;

and are independent variables;

, , and are regression coefficients.

Finding the values of the coefficients (β) that minimize the difference between the values predicted by the linear equation and the dependent variable's actual values is the first step in fitting a linear regression model. Ordinary Least Squares (OLS) is a commonly used approach for minimizing the sum of squared discrepancies between the observed and predicted values. For port emissions, Linear Regression provides a baseline model to gauge the direct proportional impact that different predictor variables have on emission levels.

***Polynomial Regression:***

Regression analysis that models the connection between the independent variables (predictors) and the dependent variable (outcome) as a non-linear function is known as non-linear regression. Non-linear regression can handle more complex relationships than linear regression, which assumes a straight line relationship.

Objective: Modeling the relationship between one or more independent variables (X) and a continuous dependent variable (Y) using a non-linear equation is the primary objective of non-linear regression. This equation can take various forms, depending on the nature of the relationship being modeled.

=

Here, the presence of the term makes the relationship non-linear.

For port emissions, Polynomial Regression can model the nonlinear relationship between the features and the response variable, capturing more complex patterns than simple linear regression.

***Ridge Regression:***

A regularization term is added to linear regression to create ridge regression. Complex models are discouraged by this regularization term (L2 penalty), which penalizes the square of the coefficients and essentially shrinks them toward zero. This method is especially useful when dealing with multicollinearity or when the number of predictors exceeds the number of observations.

When modeling HC emissions from ships, Ridge Regression can be particularly effective because it can handle multiple correlated predictors, such as engine size, fuel type, and ship age, without suffering from the overfitting that can plague standard linear regression. This produces a model that is resilient for predictive analytics in environmental research, performing well on training data and generalizing well to new, unseen data.

***XGBoost Regressor:***

A gradient boosting structure serves as the basis for the machine learning method referred to as XGBoost Regressor, which has garnered significant attention because of its high performance on a variety of tasks including regression. XGBoost acronym means eXtreme Gradient Boosting, and it's a powerful algorithm that allows handling high-dimensional feature spaces and large datasets better.

XGBoost can use both parallel computing and distributed processing power, which are responsible for accelerating the training speed. Along with these regularization techniques that prevent overfitting due to the noise in the input data, it also incorporates generalization techniques. The algorithm offers flexibility in parameterization and design of the internal loss function and objective function for solving various regression problems.

XGBoost provides a more efficient and effective implementation of Gradient Boosting, which is valuable in predicting port emissions with high accuracy and speed, especially when dealing with large and complex data.

***Multiple Linear Regression with Interaction Effects:***

A statistical technique called multiple linear regression with interaction effects takes into account not only how each independent variable affects the dependent variable on its own, but also how these factors may influence each other's relationship to the outcome. This approach is especially helpful in situations when the amount of one independent variable affects the effect of another on the dependent variable.

A multiple linear regression model with interaction terms has the following general form:

Where,

* is the dependent variable.
* … are the independent variables.
* is the y-intercept.
* … are the coefficients of the independent variables.
* represents the interaction term between variables.
* is the error term

**Explanation of metrics used for evaluation**

In the process of developing a machine learning model to predict hydrocarbon (HC) emissions from ships at ports, the evaluation phase is crucial to ascertain the performance and applicability of the model. This phase, often termed model evaluation or assessment, provides insights into the model's accuracy, identifying both its strengths and potential limitations through a variety of metrics.

***Explained Variance Score:***

This score reflects the degree to which our model accounts for the variation observed in the dataset. It is calculated as:

An explained variance scores close to 1 indicates that our model has a performance nearly identical to the true data, with a lower value pointing to a model that may be missing key explanatory factors.

***Mean Absolute Error (MAE):***

The absolute variation between the "true" value and the measured value is represented by the term "absolute error." Mean Absolute Error aggregates these individual absolute errors across the dataset and averages them, providing a straightforward metric of average error magnitude:

In terms of target variable prediction, a lower mean absolute error (MAE) indicates a higher precision of the model.

***Mean Squared Error (MSE):***

The estimated and actual values' average squared difference is what the Mean Squared Error (MSE) quantifies. This metric places more emphasis on larger errors due to the squaring of each term:

***R Squared (R²):***

The amount of variance in the dependent variable that can be anticipated from the independent variable(s) is measured by *R²*, also known as the coefficient of determination. It is defined as:

Here, RSS (Residual Sum of Squares) represents the variance that the model fails to capture, and TSS (Total Sum of Squares) is the total variance in the data. An R² value closer to 1 indicates a model that accurately reflects the observed outcomes.

These evaluation metrics, when applied to the HC emissions data, enable us to understand the efficacy of our model. They collectively offer a comprehensive picture of how well the model predicts emissions, which in turn informs further refinement and application of the model in environmental policy and decision-making within the maritime industry.

# Results and Analysis

# Our objective is to predict the Total Emissions, a continuous variable, making this a regression problem. We aim to model the relationship between various features related to ships and their corresponding emission levels. The objective was to develop a model that, without overfitting the data, produced the lowest MAE/MSE values and the best value.

Model Performance Overview

The table shows the metrics evaluation of all the models:

**Table 1**

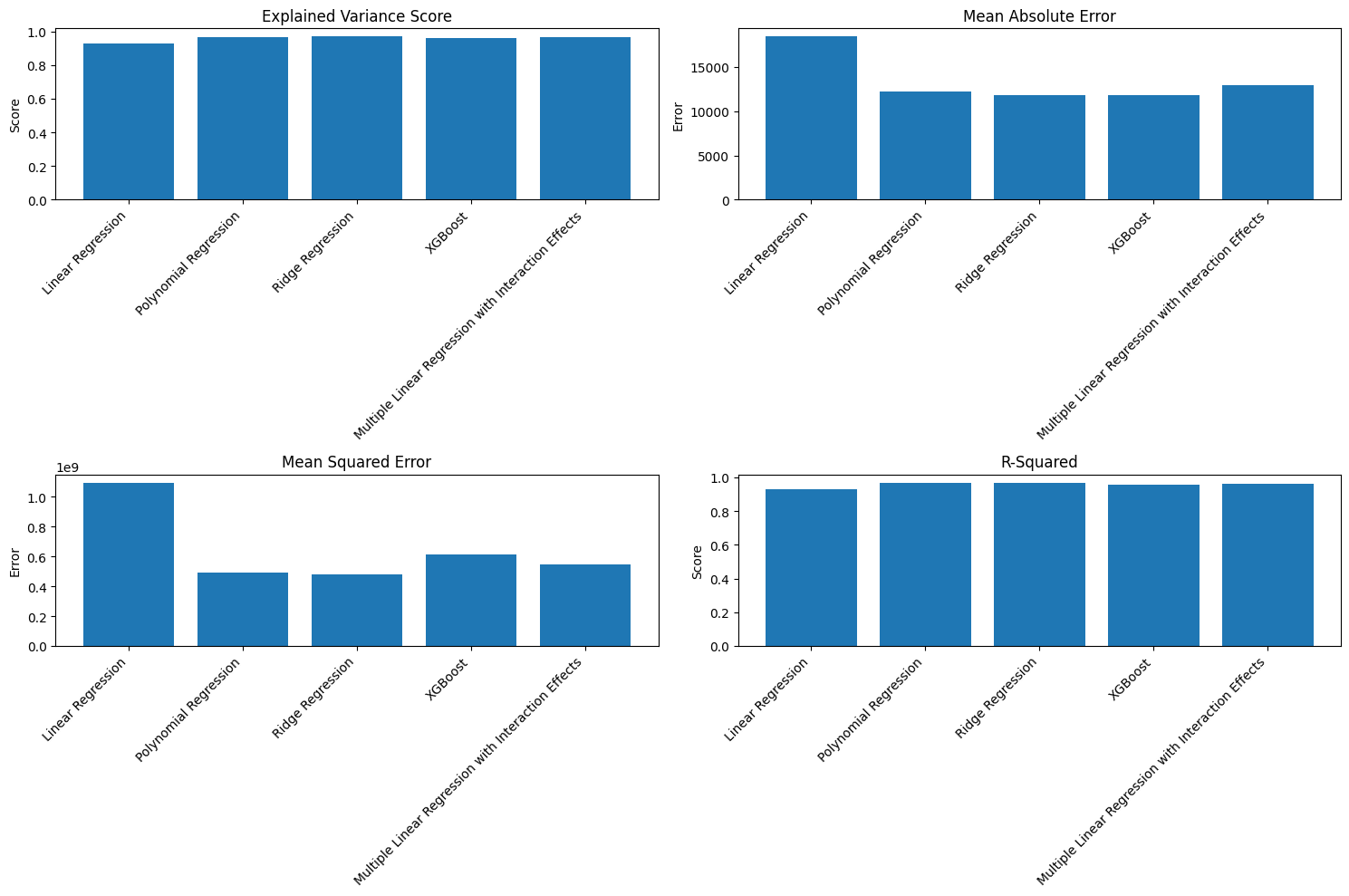
*Metrics evaluation of all the models*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Rank** | **Model** | **R² (R-squared)** | **MSE (Mean Squared Error)** | **MAE (Mean Absolute Error)** | **Explained Variance** |
| **1** | **Ridge Regression** | 0.9686 | 476,888,724.11 | 11,822.40 | 0.9686 |
| **2** | **Polynomial Regression** | 0.9674 | 494,494,792.82 | 12,221.71 | 0.9674 |
| **3** | **Multiple Linear Regression with Interaction Effects** | 0.9638 | 548,607,161.95 | 12,896.29 | 0.9639 |
| **4** | **XGBoost Regressor** | 0.9594 | 616,542,545.97 | 11,774.47 | 0.9594 |
| **5** | **Linear Regression** | 0.9279 | 1,094,740,295.03 | 18,430.01 | 0.9281 |

Our analysis employed several predictive models to estimate HC emissions, yielding varied performance across metrics. Ridge Regression demonstrated the highest accuracy, with an R-squared value of 0.9686, closely followed by Polynomial Regression (R² = 0.9674) and Multiple Linear Regression with Interaction Effects (R² = 0.9638). XGBoost, while highly competitive, ranked slightly lower (R² = 0.9532). Linear Regression, despite its simplicity, showed commendable performance with an R² of 0.9279.

**Figure 5**

*Comparison between different evaluation metrics*



In the comprehensive analysis of models to predict hydrocarbon (HC) emissions from ships at ports, the Ridge Regression model emerged as the standout performer, striking an optimal balance between model complexity and predictive accuracy. This model, distinguished by its incorporation of L2 regularization, adeptly handles multicollinearity among predictors and mitigates the risk of overfitting, a crucial advantage when dealing with the multifaceted nature of HC emissions data. Ridge Regression, which has the lowest Mean Squared Error of all the models examined and a high R-squared value of 0.9686, penalizes big coefficients to maintain the model's robustness and generalizability across various data segments. The efficacy of Ridge Regression in this context underlines the importance of selecting models that not only fit the training data well but also possess the ability to generalize to unseen data. This model’s ability to accurately capture the variance in HC emissions and minimize prediction errors provides vital insights into the factors influencing emissions. It identifies key variables impacting HC emissions, such as fuel type, engine efficiency, and ship operational practices, offering a quantitative foundation for targeted emission reduction strategies.

While Ridge Regression stood out for its overall performance, other models also demonstrated valuable characteristics in the prediction task. Polynomial Regression, for instance, showcased its strength in modeling non-linear relationships, achieving a noteworthy R-squared value. Its approach to incorporating polynomial terms allowed for a nuanced understanding of the complex dynamics influencing emissions. However, the potential for overfitting and the increased computational complexity compared to Ridge Regression were considerations that favored the latter.

Similarly, the XGBoost Regressor offered powerful predictive capabilities, particularly in handling complex and non-linear relationships with a high degree of flexibility. Its advanced algorithmic approach and feature importance insights were instrumental in understanding the multifaceted factors driving HC emissions. Nonetheless, the Ridge Regression model's superior balance between accuracy and model simplicity, coupled with its lower susceptibility to overfitting, ultimately made it the preferred choice.

Multiple Linear Regression with Interaction Effects also provided significant insights, especially into how variables interact to impact emissions. This model illuminated the combined effects of different operational and environmental factors, underscoring the complexity of HC emissions. Yet, when it came to generalizability and minimizing prediction errors, Ridge Regression maintained its lead.

**Managerial Insights**

The nuanced analysis provided by the Ridge Regression model, complemented by insights from other predictive models, lays a robust foundation for strategic decision-making aimed at reducing HC emissions in maritime operations. This section distills key insights into actionable strategies, emphasizing a data-driven approach to enhancing environmental sustainability within the maritime sector.

A critical insight from our analysis is the substantial impact of fuel type on emissions. Decision-makers are thus urged to prioritize the shift towards cleaner, low-sulfur fuels and explore the potential of LNG and biofuels. This shift not only aligns with evolving environmental regulations but also positions companies as leaders in sustainability. Complementing this fuel strategy, investments in fuel efficiency technologies ranging from advanced engine designs to air lubrication systems emerge as essential. These technologies not only reduce emissions but also optimize operational costs, presenting a compelling case for their adoption.

Operational practices, particularly those related to ship speed and routing, significantly influence emissions. The adoption of just-in-time (JIT) arrival systems, coupled with AI-enhanced route optimization, presents an opportunity to minimize fuel consumption and reduce emissions. Furthermore, the insights underscore the environmental impact of port operations, particularly during docking. Investment in onshore power supplies to enable ships to switch off their engines while berthed is recognized as a crucial technique for lowering port emissions.

Navigating the regulatory landscape requires a proactive stance. Leveraging predictive insights to anticipate and adapt to regulatory changes can minimize compliance risks and capitalize on early compliance incentives. In parallel, adopting transparent reporting practices on emissions reduction efforts not only enhances a company's environmental credentials but also sets industry benchmarks, encouraging a culture of continuous improvement.

The path to sustainable maritime operations is a collective journey. Engaging in industry consortia offers a platform for sharing best practices and driving joint initiatives in green technology. Collaboration with regulatory bodies can also facilitate the development of practical and effective environmental regulations. Furthermore, engaging customers in discussions about the environmental impact of shipping choices can enhance brand loyalty and promote a broader commitment to sustainability.

At the core of these strategies is the utilization of data. The insights from predictive modeling underscore the value of a data-driven approach in identifying emission reduction opportunities and optimizing operations. Building capabilities in data analytics enables real-time monitoring and decision-making, enhancing operational agility and strategic planning.

**Conclusion**

This research journey into understanding and predicting hydrocarbon (HC) emissions from maritime operations has culminated in significant findings, with the analysis of various predictive models lying at its core. Among the models explored Ridge Regression emerged as notably effective, striking a balance between accuracy and the complexity of capturing emissions data. This model, alongside insights gleaned from Polynomial Regression, XGBoost, and other methodologies, illuminated the key factors influencing HC emissions, including the critical role of fuel types, operational efficiencies, and the potential of technological advancements.

The study produced practical recommendations for reducing hydrocarbon emissions, with a focus on switching to cleaner fuels and streamlining ship operations using sophisticated routing and speed control. These methods support the growing regulatory requirements for sustainability in the maritime sector while simultaneously attempting to lessen the environmental impact of shipping operations.

Furthermore, this study highlighted the importance of collaborative efforts among maritime stakeholders. It suggests that sharing knowledge, engaging in industry-wide sustainability initiatives, and adopting transparent reporting practices are essential steps toward achieving significant emissions reductions.

In concluding this analysis, it's evident that predictive modeling offers a powerful tool for understanding and addressing the environmental challenges faced by the maritime sector. By leveraging data-driven insights, the industry can navigate towards more sustainable practices, reducing HC emissions and contributing to global environmental conservation efforts.

In reflection, this research journey has broadened the understanding of the factors driving HC emissions in maritime operations and underscored the transformative potential of predictive analytics in shaping a more sustainable future for the maritime industry.

# 

# References

*Volatile organic compounds’ impact on indoor air quality | US EPA*. (2023, August 15). US EPA. <https://www.epa.gov/indoor-air-quality-iaq/volatile-organic-compounds-impact-indoor-air-quality>

Deniz, C., Kılıç, A., & Cıvkaroglu, G. (2010). Estimation of shipping emissions in Candarli Gulf, Turkey. *Environmental Monitoring and Assessment*, *171*(1–4), 219–228. <https://doi.org/10.1007/s10661-009-1273-2>

Wu, H., Wang, C., Chen, E., & Ye, Z. (2023). Development of a spectrum-based ship fuel sulfur content real-time evaluation method. *Marine Pollution Bulletin*, *188*, 114484. <https://doi.org/10.1016/j.marpolbul.2022.114484>

Aakko-Saksa, P., Lehtoranta, K., Kuittinen, N., Järvinen, A., Jalkanen, J., Johnson, K. C., Jung, H., Ntziachristos, L., Gagné, S., Takahashi, C., Karjalainen, P., Rönkkö, T., & Timonen, H. (2023). Reduction in greenhouse gas and other emissions from ship engines: Current trends and future options. *Progress in Energy and Combustion Science*, *94*, 101055. <https://doi.org/10.1016/j.pecs.2022.101055>

Azarkamand, S., Wooldridge, C., & Darbra, R. M. (2020). Review of initiatives and methodologies to reduce CO2 emissions and climate change effects in ports. *International Journal of Environmental Research and Public Health*, *17*(11), 3858. <https://doi.org/10.3390/ijerph17113858>

*Air pollution from ships or ports harboring a problem?* (n.d.). Aeroqual. <https://www.aeroqual.com/blog/ship-pollution-port-air-quality>

Yeh, C., Lin, C., Shen, H., Cheruiyot, N. K., Nguyen, D., & Chang, C. (2022). Real-time energy consumption and air pollution emission during the transpacific crossing of a container ship. *Scientific Reports*, *12*(1). <https://doi.org/10.1038/s41598-022-19605-7>

Wan, Z., Zhang, T., Sha, M., Guo, W., Jin, Y., Guo, J., & Liu, Y. (2021). Evaluation of emission reduction strategies for berthing containerships: A case study of the Shekou Container Terminal. *Journal of Cleaner Production*, *299*, 126820. <https://doi.org/10.1016/j.jclepro.2021.126820>

Bjerkan, K. Y., & Seter, H. (2019). Reviewing tools and technologies for sustainable ports: Does research enable decision making in ports? *Transportation Research Part D: Transport and Environment*, *72*, 243–260. <https://doi.org/10.1016/j.trd.2019.05.003>

Cammin, P., Yu, J., & Voß, S. (2022). Tiered prediction models for port vessel emissions inventories. *Flexible Services and Manufacturing Journal*, *35*(1), 142–169. <https://doi.org/10.1007/s10696-022-09468-5>

Yao, Y., Sun, R., Sun, Y., & Shu, Y. (2022). Integrated Carbon Emission Estimation Method and Energy Conservation Analysis: The Port of Los Angles Case Study. *Journal of Marine Science and Engineering*, *10*(6), 717. <https://doi.org/10.3390/jmse10060717>

Xiao, G., Wang, T., Luo, Y., & Yang, D. (2023). Analysis of port pollutant emission characteristics in United States based on multiscale geographically weighted regression. *Frontiers in Marine Science*, *10*. <https://doi.org/10.3389/fmars.2023.1131948>

Flowers, Daniel L, et al. “Prediction of Carbon Monoxide and Hydrocarbon Emissions in Iso-Octane HCCI Engine Combustion Using Multizone Simulations.” *Proceedings of the Combustion Institute*, vol. 29, no. 1, 1 Jan. 2002, pp. 687–694, <https://doi.org/10.1016/s1540-7489(02)80088-8>. Accessed 27 Mar. 2024.

Fu, X., Chen, D., Guo, X., Lang, J., & Zhou, Y. (2022). Improving the estimation of ship emissions using the high‐spatiotemporal resolution wind fields simulated by the Weather Research and Forecast model: A case study in China. *Journal of Industrial Ecology*, *26*(6), 1871–1881. <https://doi.org/10.1111/jiec.13278>

Paternina-Arboleda, C. D., Agudelo-Castañeda, D. M., Voß, S., & Das, S. (2023). Towards cleaner ports: Predictive modeling of sulfur dioxide shipping emissions in maritime facilities using machine learning. *Sustainability*, *15*(16), 12171. <https://doi.org/10.3390/su151612171>

Hussaın, I., Wang, H., Muhammad, S., Ho, Q. B., Wemegah, T. D., & Noor, S. (2022). Estimation of shipping emissions in developing country: A case study of Mohammad bin Qasim Port, Pakistan. *International Journal of Environmental Research and Public Health*, *19*(19), 11868. <https://doi.org/10.3390/ijerph191911868>

Port of San Diego. (2013). *Port of San Diego Climate Action Plan 2013*. <https://pantheonstorage.blob.core.windows.net/environment/Port-of-San-Diego-Climate-Action-Plan.pdf>

Zhang, Y., Peng, Y., Wang, W., Gu, J., Wu, X., & Feng, X. (2017). Air emission inventory of container ports’ cargo handling equipment with activity-based “bottom-up” method. *Advances in Mechanical Engineering (Print)*, *9*(7), 168781401771138. <https://doi.org/10.1177/1687814017711389>

Darbra, R. M., Ronza, A., Casal, J., Stojanovic, T., & Wooldridge, C. (2004). The self-diagnosis method. *Marine Pollution Bulletin*, *48*(5–6), 420–428. <https://doi.org/10.1016/j.marpolbul.2003.10.023>

Wright, P. (2013). Impacts of climate change on ports and shipping. MCCIP Sci. Rev, 2013, 263-270. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Impacts+of+climate+change+on+ports+and+shipping&author=Wright,+P.&publication_year=2013&journal=MCCIP+Sci.+Rev.&pages=263%E2%80%93270&doi=10.14465/2013.arc28.263-270)]

Henein, N. A., & Tagomori, M. K. (1999). Cold-start hydrocarbon emissions in port-injected gasoline engines. *Progress in Energy and Combustion Science*, *25*(6), 563–593. <https://doi.org/10.1016/s0360-1285(99)00003-9>

Lang, K. R., & Cheng, W. K. (2006). Effects of fuel injection strategy on HC emissions in a Port-Fuel-Injection engine during fast idle. *SAE Technical Paper Series*. <https://doi.org/10.4271/2006-01-3400> Merk, O. & International Transport Forum. (2014).

Shipping emissions in ports. In *Discussion Paper*. <https://www.itf-oecd.org/sites/default/files/docs/dp201420.pdf>

Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and Health Impacts of Air Pollution: A review. *Frontiers in Public Health*, *8*. <https://doi.org/10.3389/fpubh.2020.00014>

Lee, H., Pham, H. T., Chen, M., & Choo, S. (2021). Bottom-Up approach ship emission inventory in Port of Incheon based on VTS data. *Journal of Advanced Transportation*, *2021*, 1–16. <https://doi.org/10.1155/2021/5568777>

Karakikes, I., Hofmann, W., Mitropoulos, L., & Savrasovs, M. (2018). Evaluation of freight measures by integrating simulation tools: the case of Volos Port, Greece. *Transport and Telecommunication (Print)*, *19*(3), 224–232. <https://doi.org/10.2478/ttj-2018-0019>

Gonzalez-Aregall, M., & Bergqvist, R. (2020). Green port initiatives for a more sustainable port-city interaction: The case study of Barcelona. In *Elsevier eBooks* (pp. 109–132). <https://doi.org/10.1016/b978-0-12-819134-7.00007-1>

Barberi, S., Sambito, M., Neduzha, L. O., & Severino, A. (2021). Pollutant Emissions in Ports: A Comprehensive review. *Infrastructures*, *6*(8), 114. <https://doi.org/10.3390/infrastructures6080114>

U.S. Environmental Protection Agency. (2016). National Port Strategy Assessment: Reducing Air Pollution and Greenhouse Gases at U.S. ports. In *Office of Transportation Air Quality* [Report]. <https://19january2017snapshot.epa.gov/sites/production/files/2016-09/documents/420s16002.pdf>

Environmental and Energy Study Institute (EESI). (n.d.). *Issue Brief | Climate Change Mitigation and Adaptation at U.S. Ports (2022) | White Papers | EESI*. <https://www.eesi.org/papers/view/issue-brief-climate-change-mitigation-and-adaptation-at-u.s-ports-2022>

Bojić, F., Gudelj, A., & Bošnjak, R. (2022). Port-Related Shipping Gas Emissions—A Systematic Review of Research. *Applied Sciences*, *12*(7), 3603. <https://doi.org/10.3390/app12073603>

*Ports Initiative | US EPA*. (2024, March 25). US EPA. <https://www.epa.gov/ports-initiative>

## Acknowledgments

I would like to acknowledge the use of OpenAI's ChatGPT for assistance in refining the English language and structuring certain sections of this report.